پژوهش‌های نوین در تصمیم‌گیری
دوره ۱۰، شماره ۵، زمستان ۱٤۰٤، صص ٦۱-۹۰
نوع مقاله: پژوهشی

طراحی مدل هوشمند اعتباریابی واردکنندگان کالا با رویکرد یادگیری ماشین

سید سینا مدنی [۱]، محمود دهقان نیری [۲*]، علی رجب زاده قطری [۳]

۱. دانشجوی دکتری مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
۲. دانشیار، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران
۳. استاد، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه تربیت مدرس، تهران، ایران

تاریخ دریافت: ۱٤۰۳/۱۱/۳۰ تاریخ پذیرش: ۱٤۰٤/۰۵/۰٦

چکیده

تفاوت نرخ ارز رسمی با نرخ ارز در بازار آزاد زمینه سوءاستفاده اخلالگران ارزی را فراهم نموده است. از سویی دیگر واردکنندگان کالا بدون توجه به سوابق عملکرد ارزی و ریالی، مورد ارزیابی شبکه بانکی کشور قرار گرفته و نسبت به اخذ وثایق ریالی از ایشان اقدام می‌شود. هدف اصلی این پژوهش طراحی مدل هوشمند عمیق به منظور ارزیابی ریسک و اخذ وثایق از واردکنندگان کالا، به نحوی است که مدل نهایی با بالاترین میزان صحت و دقت، توانایی تحلیل داده‌های بزرگ عملکردی را داشته باشد. بدین منظور، در ابتدا داده‌های عملکردی با استفاده از روش کی‌مینز خوشه‌بندی شده و سپس نتایج مدل خوشه‌بندی به عنوان ورودی مدل‌های طبقه‌بندی شامل روش‌های جنگل تصادفی، اکس جی بوست و شبکه عصبی کراس ترتیبی مورد استفاده قرار گرفته و نتایج هر مدل با استفاده از شاخص F مورد بررسی قرار گرفته‌اند. در نهایت، مدل ترکیبی خوشه‌بندی کی‌مینز- شبکه عصبی کراس ترتیبی به عنوان مدلی که بالاترین میزان شاخص صحت و دقت را داشته انتخاب و با استفاده از مدل یاد شده مشتریان در ۳ طبقه قرار گرفته و ترکیب وثایق نقد و غیر نقد برای ایشان تعیین شده است. نتایج بیانگر آن است که مدل نهایی به خوبی توانایی طبقه‌بندی مشتریان را با توجه به سوابق عملکردی ایشان، داشته است.

کلیدواژه‌ها: شبکه عصبی عمیق، ارزیابی ریسک، هوش مصنوعی، یادگیری ماشین

---

* نویسنده مسئول مقاله E-mail: mdnayeri@modares.ac.ir

۱- مقدمه و بیان مسئله

تجارت خارجی یا به عبارتی واردات و صادرات کالا و خدمات، یکی از بخش‌های مهم اقتصاد هر کشور بوده، به نحوی که گسترش تجارت خارجی نقشی تاثیرگذار بر بهبود شاخص‌های عملکردی اقتصادی کشور دارد [۱]. در این بین "واردات کالا و خدمات" ضامن بقای تولید داخلی و برآورده کننده بخشی از نیازهای بازار است که توسط تولیدکنندگان داخلی تامین نمی‌شوند. شایان ذکر است که بر اساس آمار سازمان توسعه تجارت، مجموع تجارت خارجی کشور در سال ۱٤۰۲ رقمی بالغ بر ۱۱۵ میلیارد دلار آمریکا بوده که از این مبلغ سهم واردات کالا رقمی بالغ بر ٦٦ میلیارد دلار آمریکا بوده است [۲]. طبق بند ۱٤ مقررات ارزی بانک مرکزی، پیش از تأمین ارز، مشتری موظف به ارائه وثیقه ریالی معادل ۳۵ درصد مبلغ ارز به عنوان تضمین ورود کالا و ارائه پروانه گمرکی است. در صورت عدم رفع تعهد ارزی در مهلت مقرر، بانک پس از یک سال وثایق را اجرا کرده و وجوه حاصل را به حساب‌های تعیین‌شده توسط بانک مرکزی واریز می‌کند [۳]. هرچند تعیین ترکیب وثایق بر عهده بانک‌هاست، اما اکثر بانک‌ها برای جذب مشتری و بدون توجه به وضعیت اعتباری، صرفاً از اسناد تجاری (به‌ویژه سفته) استفاده می‌کنند که زمینه سوءاستفاده اخلال‌گران ارزی را فراهم کرده است [٤]. اختلاف نرخ ارز رسمی با نرخ بازار آزاد (حدود ۳۰ درصد) و اختلاف بسیار بیشتر با سایر نرخ‌ها (مانند نرخ ترجیحی) انگیزه سوءاستفاده از دریافت ارز رسمی و فروش آن در بازار آزاد را افزایش داده است [۵].

از سویی دیگر، در تجارت خارجی روش‌های اصلی پرداخت شامل حواله ارزی (پیش‌پرداخت کامل)، وصولی اسنادی و اعتبار اسنادی است. تفاوت کلیدی در نحوه پرداخت وجه به فروشنده و ریسک طرفین است. در روش حواله ارزی، واردکننده ابتدا تمام وجه را پرداخت می‌کند و سپس فروشنده کالا را ارسال می‌کند که این امر ریسک بالایی برای واردکننده ایرانی و خطر خروج منابع ارزی بدون ورود کالا به همراه دارد [٦]. با این حال، به دلیل شرایط ارزی کشور، فروشندگان خارجی از پذیرش روش‌های اعتبار اسنادی و وصولی اسنادی (که ریسک کمتری برای واردکننده دارند) خودداری کرده و تنها با روش حواله ارزی (پُرریسک‌ترین روش برای واردکننده) اقدام به ارسال کالا می‌کنند که این موضوع ریسک و هزینه‌های واردکنندگان ایرانی را افزایش می‌دهد [۷].

هدف این پژوهش، ارزیابی ریسک واردکنندگان کالا و اخذ وثایق متناسب با ریسک هر یک، به منظور جلوگیری از تخصیص و فروش ارز به اخلالگران ارزی، مدیریت منابع ارزی کشور و پوشش ریسک ناشی از روش پرداخت حواله ارزی است. با توجه به تعداد زیاد شاخص‌های تأثیرگذار و تعدد واردکنندگان، استفاده از مدل‌های سنتی توسط بانک مرکزی هزینه‌بر و مستعد اعمال سلایق فردی خواهد بود. بنابراین، استفاده از رویکردهای نوین نظیر یادگیری ماشین[۱] و فناوری اطلاعات راهکار مناسبی است. هدف اصلی مقاله، طراحی مدلی هوشمند برای ارزیابی ریسک واردکنندگان کالا ویژه بانک مرکزی ج.ا.ا است تا بتواند حجم زیادی از مشتریان ارزی را از نظر ویژگی‌های ارزی و ریالی تحلیل کند.

درنهایت با استفاده از مدل نهایی، بانک مرکزی می‌تواند نسبت به تعیین ریسک واردکنندگان کالا اقدام نموده و بانک‌های عامل را موظف نماید تا با توجه به ریسک هر واردکننده نسبت به تعیین ترکیب وثایق اقدام نمایند تا ضمن اطمینان از تخصیص ارز به واردکنندگان واقعی و درنتیجه ورود کالا به کشور، امکان سوء استفاده اخلال‌گران به حداقل ممکن کاهش یابد. مهمترین تفاوت مدل نهایی با مدل‌های پیشین، استفاده از روش‌های مبتنی بر یادگیری عمیق بوده، که علت آن ماهیت داده‌های کلانِ مورد استفاده در پژوهش است.

۲- ادبیات و پیشینه پژوهش

ریسک اعتباری به معنای احتمال از دست دادن اصل سرمایه یا سود مورد انتظار به دلیل عدم بازپرداخت وام یا عدم انجام تعهدات قراردادی از سوی وام‌گیرنده است [۸]. در سال‌های اخیر، ارزیابی و مدیریت این ریسک همواره مورد توجه بوده و محققان با رویکردهای مختلفی از جمله روش‌های آماری، روش‌های مبتنی بر هوش مصنوعی و روش‌های ترکیبی به ارائه راهکار پرداخته‌اند. سنجش ریسک اعتباری در داخل کشور نیز توسط پژوهشگران ایرانی مورد توجه قرار گرفته است. در ادامه این روش‌ها بررسی می‌شوند.

روش‌های آماری و سنتی: روش‌های آماری همواره به منظور سنجش ریسک اعتباری مورد نظر پژوهشگران بوده است. اِمل[۲] و همکاران در سال ۲۰۰۳ از تحلیل پوششی داده‌ها[۳] برای مدیریت ریسک و رتبه‌بندی اعتباری ۸۲ شرکت متقاضی دریافت تسهیلات از بانک‌های تجاری

---

۱ Machine Learning
۲ Emmel
۳ Data Envelopment Analaysis

ترکیه اقدام و داده‌های مرتبط با نسبت های مالی را برای طیف مختلفی از شرکت‌های دریافت کننده تسهیلات شامل شرکت‌های فعال در حوزه وسایل الکتریکی، صنایع غذایی، شرکت‌های شیمیایی فلزی و حمل و نقل در حوزه چرم و پارچه مورد بررسی قرار داده‌اند. با توجه به بررسی صورت گرفته توسط ایشان، تا دهه ابتدایی سال ۲۰۰۰ بیشتر پژوهش‌های صورت گرفته در خصوص مدیریت ریسک، بر روش‌های آماری و روش‌های مبتنی بر "پژوهش در عملیات" تکیه کرده‌اند [۹].

روش‌های مبتنی بر یادگیری ماشین: ابتدای قرن جاری را می‌توان نقطه شروع استفاده از روش‌های مبتنی بر هوش مصنوعی برای مدیریت ریسک در نظر گرفت. به نحوی که در ابتدا وست[۱] در سال ۲۰۰۰ نسبت به مقایسه ۲ مدل پارامتریک، ۳ مدل ناپارامتریک و ۵ ساختار معماری شبکه عصبی اقدام نموده است. در پژوهش یاد شده ۲ مجموعه داده اطلاعات اعتباری آلمان و استرالیا با توجه به متغیرهایی ازجمله اطلاعات هویتی مانند سن، جنسیت، شغل، وضعیت مسکن، ملیت، وضعیت تاهل و اطلاعات اعتباری مانند تاریخچه اعتبار، عملکرد حساب، هدف از اخذ تسهیلات و مقدار تسهیلات استفاده شده‌اند. پس از مقایسه ۱۰ مدل با توجه به شاخص صحت، مشخص گردید که مدل شبکه عصبی ترکیب متخصصان[۲] بهترین میزان صحت را داشته است. همچنین در بین مدل‌های پارامتریک و ناپامتریک، مدل رگرسیون لجستیک[۳] بهترین میزان صحت را داشته است [۱۰]. پس از ایشان، توانا[۴] و همکارانش در سال ۲۰۱۸ از شبکه‌های عصبی مصنوعی[۵] به منظور ارزیابی ریسک نقدینگی در صنعت بانکداری بهره برده‌اند. بدین منظور ایشان نسبت به تعریف ۱۰ شاخص بر اساس ویژگی‌هایی همچون مجموع تسهیلات، مجموع سپرده، دارایی‌های نقد و مانند آن اقدام نموده‌اند [۱۱]. ونسچی[۶] و همکارانش در سال ۲۰۱۸ با استفاده از برنامه‌ریزی ژنتیک[۷] و یادگیری ماشین مدلی را برای پیش‌بینی ریسک مشتریان بانک در عدم پرداخت اقساط تسهیلات بانکی ارائه نموده‌اند [۱۲]. بدین منظور داده‌های مالی مرتبط با ٥٦٦٦٩ درخواست مورد بررسی و لیبل‌زنی (به عنوان مشتری خوب و بد) قرار گرفته و با بهره‌گیری از داده‌های یاد شده،۱۹ متغیر طراحی شده است. از متغیرهای

---

[۱] West
[۲] Mixture-of-experts neural network
[۳] Logistic regression
[۴] Tavana
[۵] Artificial Neural Networks
[۶] Vanneschi
[۷] Genetic Programming

یاد شده می‌توان به متغیرهای مرتبط با محل فعالیت مشتری، عملکرد مالی، میانگین تسهیلات اشاره نمود.

یکی از نکاتی که در بکارگیری یادگیری ماشین و هوش مصنوعی در مدیریت ریسک همواره مدنظر محققان بوده، چگونگی انتخاب ویژگی به منظور دسته‌بندی و یا خوشه‌بندی است. بدین منظور اورسکی[۱] و همکارانش در سال ۲۰۱٤ با بررسی داده‌های مشتریان حقیقی شامل داده-های هویتی (سن، جنسیت، شغل)، داده‌های مالی (تعداد و مجموع گردش مالی، میانگین گردش حساب، تعداد و میانگین چک‌های برگشتی)، داده‌های تسهیلاتی (تاریخچه تسهیلات دریافتی، تعداد و میانگین تسهیلات دریافتی) نسبت به ارائه مدلی برای انتخاب ویژگی با استفاده از الگوریتم ژنتیک[۲] اقدام نموده‌اند [۱۳].

تای[۳] و همکاران در سال ۲۰۱۹ نسبت به بکارگیری ۲ مدل شبکه عصبی یادگیری عمیق شامل شبکه عصبی متوالی عمیق[٤] و شبکه عصبی پیچشی[٥] اقدام نموده‌اند. در این پژوهش ۳ مجموعه داده مورد استفاده قرار گرفته شده که عبارتند از مجموعه داده اعتباری آلمان شامل ۱۰۰۰ رکورد و ۲٤ متغیر، مجموعه داده اعتباری استرالیا شامل ٦۹۰ رکورد و ۱٤ متغیر و مجموعه داده اعتباری ککل[٦] گوگل شامل ۲٥۰۰۰۰ رکورد و ۱۰ متغیر. در نهایت با توجه به شاخص‌های صحت، دقت، فراخوانی و شاخص F۱ ایشان بیان نمودند که در هنگام سنجش رتبه اعتباری، مدل‌های شبکه عصبی عمیق عملکرد بهتری نسبت به مدل‌های شبکه عصبی ساده دارند. درنهایت پیشنهاداتی به منظور بهبود عملکرد مدل‌های عمیق توسط ایشان ارائه شده است [۱٤]. منگ[۷] و همکاران در همان سال نسبت بررسی داده‌های تاریخی تسهیلات پرداختی موسسات مالی اقدام نموده با استفاده از روش جنگل تصادفی[۸] نسبت به ارائه مدل مدیریت ریسک اقدام نموده اند. در پژوهش یاد شده، با بررسی ۱٤ متغیر مرتبط با داده‌های مالی برای بازه زمانی ۳ ساله، پیش‌بینی احتمال شکست تسهیلات، مبنایی برای کنترل ریسک در نظر گرفته شده است [۱٥]. همچنین ژائو[۹] و همکاران با ترکیب روش تاپسیس[۱۰] با روش جنگل تصادفی، مدلی را جهت مدیریت ریسک در موسسات مالی بسط داده‌اند. در پژوهش یاد شده

---

[۱] Oreski
[۲] Genetic algorithm
[۳] Tai
[٤] Sequential Deep Neural Network
[٥] Convolutional Neural Network
[٦] kaggle
[۷] Meng
[۸] Random Forest
[۹] Zhao
[۱۰] TOPSIS

نسبت به وزن‌دهی ۷ متغیر شامل میانگین درآمد سالانه، میانگین فروش سالانه، میانگین نرخ رشد سالانه سود، میانگین نرخ رشد سالانه درآمد، نرخ تبادلات غیر معتبر، بالاترین حد نوسانات و نوسانات درآمد اقدام و در نهایت رتبه اعتباری برای ۱۲۳ شرکت مشخص شده است [۱٦].

از سویی دیگر وانگ[۱] و همکاران در سال ۲۰۲۲ مدل ارزیابی ریسک افراد را با استفاده از مدل اکس جی بوست[۲] طراحی نموده‌اند. در پژوهش یاد شده در ابتدا ۳ روش برای انتخاب ویژگی مورد بررسی قرار گرفته، سپس با استفاده از داده‌های عملکرد ۱۰ هزار فرد پس از بکارگیری روش‌های جنگل تصادفی، روش نزدیکترین همسایه و روش اکس جی بوست، این روش به عنوان مدل نهایی جهت ارزیابی ریسک افراد برای بانک‌ها انتخاب شده است. در این پژوهش ۲٦ ویژگی شامل داده‌های هویتی (سن، جنسیت، وضعیت تاهل) و سوابق مالی (سوابق تسهیلات، وضعیت درآمد، وضعیت بدهی خانوار) بررسی شده است. همچنین روش‌های جنگل تصادفی و اکس جی بوست به عنوان موثرترین روش برای بررسی داده‌های کلان معرفی شده‌اند [۱۷].

بررسی پیشینهٔ پژوهش بیانگر این امر است که با بهبود سیستم‌های سخت‌افزاری کامپیوتری در انتهای دهه دوم قرن جاری، یادگیری عمیق[۳] مورد توجه بیشتر محققان قرار گرفته شده است. نیاز به تحلیل داده‌های بزرگ نیز در این استقبال تاثیرگذار بوده، به نحوی که لی[٤] و همکارانش در سال ۲۰۱۷ با بررسی سوابق داده‌های مالی و تسهیلاتی مشتریان یک الگوریتم شبکه عصبی عمیق[٥] برای خوشه‌بندی[٦] و ارزیابی ریسک اعتباری استفاده نموده‌اند [۱۸]. همچنین ژو[۷] و همکارانش در سال ۲۰۱۸ با بررسی مجموعه داده مشتریان بانک‌های چینی و با بهره‌گیری از ویژگی‌هایی شامل سوابق مالی (مانده حساب، میانگین حساب) و تسهیلاتی مدل ترکیبی یادگیری عمیق را برای رتبه بندی اعتباری مشتریان معرفی نموده‌اند [۱۹]. علاوه بر موارد ذکر شده آدو[۸] و همکارانش در همان سال با استفاده از مدل‌های یادگیری عمیق و با استفاده از داده‌های مرتبط با نسبتهای مالی و ترازنامه یکصد و ده هزار شرکت، یک مدل

---

[۱] Wang
[۲] XGBoosT
[۳] Deep Learning
[٤] Li
[٥] Deep neural network
[٦] Clustering
[۷] Zhu
[۸] Addo

دسته‌بندی صفر و یک را به منظور شناسایی احتمال عدم بازپرداخت وام‌های بانکی طراحی نموده‌اند [۲۰].

در سال ۲۰۲۲ تیاگی[۱] با بررسی مدل‌های مختلف یادگیری ماشین از جمله رگرسیون لوجستیک، درخت تصمیم[۲]، جنگل تصادفی،Ada Boost و شبکه عصبی بازگشتی کراس بر اثربخشی کلی روش شبکه عصبی بازگشتی کراس به عنوان روش برتر جهت ارزیابی ریسک تاکید نمود. در این پژوهش، امکان تحلیل داده‌های بزرگ مهمترین دلیل اقبال به شبکه‌های عصبی عمیق ذکر شده، به نحویکه در پژوهش ایشان ۳۰ ویژگی مرتبط با سوابق مالی برای یکصد و ده هزار رکورد تحت بررسی قرار گرفته شده است [۲۱]. همچنین در همان سال باستوس[۳] نسبت به مقایسه مدل درخت تصمیم تقویت شده[٤] با مدل شبکه عصبی چند لایه[٥] و ماشین بردار پشتیبان[٦] اقدام نموده است. مجموعه داده‌های اعتباری آلمان و داده‌های اعتباری استرالیا توسط ایشان مورد استفاده قرار گرفته شده، که با توجه به نتایج بررسی صورت گرفته، میزان صحت مدل درخت تصمیم تقویت شده در مقایسه با دو مدل دیگر بالاتر بوده است [۲۲].

در نهایت مستیری[۷] در سال ۲۰۲٤ شش مدل رتبه‌بندی اعتباری مشتریان حقیقی بانک از جمله تحلیل تشخیصی خطی[۸]،جنگل تصادفی، رگرسیون لوجستیک، درخت تصمیم، ماشین بردار پشتیبان، و شبکه عصبی عمیق کراس را با توجه به ۳ معیار عملکردی مقایسه نموده و بر صحت و عملکرد بالاتر مدل شبکه عصبی عمیق کراس به منظور رتبه‌بندی اعتباری تاکید نموده است. ویژگی‌های مورد بررسی در این پژوهش شامل ۱۲ ویژگی از جمله سن، درآمد، وضعیت تاهل، تحصیلات، گردش حساب، میزان تسهیلات دریافتی بوده که ویژگی‌های یاد شده برای ٦۸۸ فرد حقیقی از بانک تجاری تونس اخذ شده است [۲۳].

روش‌های هیبریدی: برخی پژوهشگران به منظور ایجاد متغیر ناظر از ترکیب مدل‌های خوشه–بندی و طبقه‌بندی استفاده نموده‌اند به نحویکه در ابتدا با استفاده از روش‌های خوشه‌بندی نسبت به ایجاد متغیر ناظر اقدام شده، سپس مجموعه داده با استفاده از روش‌های طبقه بندی

---

[۱] Tyagi
[۲] Decision Tree
[۳] Bastos
[٤] Boosted decision trees
[٥] Multy layer perceptron
[٦] Support vector machines
[۷] Mestiri
[۸] Linear Discriminant Analysis

مورد بررسی و بهبود قرار گرفته‌اند. به عنوان مثال در سال ۲۰۲۰ جدوال[۱] و همکاران با بررسی سوابق هویتی (محل فعالیت، سابقه فعالیت شرکت) و سوابق مالی (مانده حساب، گردش مالی حساب) و سوابق تسهیلاتی (تعداد و مجموع تسهیلات دریافتی) نسبت به توسعه مدل هیبریدی خوشه‌بندی-طبقه‌بندی اقدام نمودند. در مدل ایشان، داده‌های آموزش در ابتدا خوشه-بندی شده و سپس به منظور کاهش خطا اقدام به استفاده از مدل‌های طبقه‌بندی شده است [۲٤].

پژوهش‌های مرتبط با ایران: پژوهشگران ایرانی مختلفی نیز در داخل کشور به بررسی ریسک اعتباری مشتریان بانک پرداخته‌اند. روند استفاده از مدل‌های مختلف در داخل کشور مشابه روند مورد استفاده پژوهشگران خارج از کشور بوده و در ابتدا روش‌های آماری و مبتنی بر تحقیق در عملیات مورد استقبال قرار گرفته‌اند، به نحویکه صفری و همکاران در سال ۱۳۸۸ نسبت به رتبه‌بندی مشتریان حقوقی بانکها برحسب ریسک اعتباری به روش تحلیل پوششی داده‌ای اقدام نموده‌اند. در پژوهش یاد شده از بین ۲۳ متغیر مالی و غیر مالی و با استفاده از تحلیل عاملی، ٦ متغیر فروش، تسهیلات بانکی، نقدینگی، فعالیت، سودآوری و بدهی انتخاب و ۱٤٦ مشتری بررسی شده‌اند[۲٥]. سپس صفایی و همکاران در سال ۱۳۹٥ با استفاده از داده-هایی شامل جنسیت، تحصیلات، محل سکونت، شغل، سابق اعتباری، وضعیت چک، نرخ سود و نوع وثیقه و با بکارگیری روش تاپسیس[۲] نسبت به طراحی مدل طبقه‌بندی مشتریان اقدام نموده و به این نتیجه رسیده‌اند که مدل‌های مبتنی بر تصمیم‌گیری چند معیاره قابلیت تحلیل بهتری نسبت به سایر مدل‌ها دارند [۲٦].

با افزایش حجم داده‌ها و استقبال پژوهشگران از روش‌های هوش مصنوعی و یادگیری ماشین در خارج از کشور، پژوهشگران داخلی نیز به تدریج از این رویکردها استفاده کرده‌اند. در سال ۱۳۹٤، حری و همکاران با استفاده از سه شاخص سن، سابقه ارتباط با بانک و میزان اعتبار، یک مدل ترکیبی عصبی-فازی را برای رتبه‌بندی مشتریان بانک طراحی و با الگوریتم مورچگان بهبود بخشیدند [۲۷]. سپس در سال ۱۳۹٥، سهرابی و همکاران با استفاده از داده‌های حساب و تسهیلات مشتریان یک بانک منتخب بزرگ، سیستم توصیه‌گر تسهیلات بانکی را طراحی کردند؛ در این پژوهش مشتریان تسهیلاتی خوشه‌بندی شده و با طبقه‌بندی، امکان پیش‌بینی

---

[۱] Jadwal

[۲] Topsis

خوشه مشتریان جدید فراهم گردید که به مدیران در تصمیم‌گیری فوری کمک می‌کرد [۲۸]. در نهایت در سال ۱٤۰۱، البرزی و همکاران با استفاده از شاخص سنجه F، عملکرد مدل‌های مختلف یادگیری ماشین شامل نزدیکترین همسایه[۱]، ماشین بردار پشتیبان و شبکه عصبی را بر اساس ویژگی‌هایی نظیر سن، مبلغ و نوع وثیقه، مبلغ و تعداد تسهیلات بخش اقتصادی برای مشتریان بانک ملت مقایسه نمودند [۲۹].

جدول شماره ۱ به بررسی و تحلیل پیشینه پژوهش‌های اشاره شده می‌پردازد. همانطور که در جدول یاد شده مشخص است، داده‌های هویتی، سوابق مالی و سوابق تسهیلاتی بیشترین نقش را در هنگام بررسی سوابق عملکردی داشته، ولیکن به دلیل تفاوت فضای تجاری کشور با فضای تجارت بین‌المللی، شاخص‌های مرتبط با عملکرد ارزی کمتر مورد توجه واقع شده‌اند. از سویی دیگر با افزایش حجم داده‌های مورد بررسی، رویکردهای نوینِ ارزیابی ریسک از جمله هوشمندسازی عمیق مورد استقبال پژوهشگران واقع شده، همچنین بررسی صورت گرفته بیانگر آن است که در صورت بکارگیری روش‌های مختلف برای یک مجموعه داده، روش‌های جنگل تصادفی، اکس جی بوست و شبکه عصبی کراس ترتیبی[۲] به عنوان روش‌های منتخب و معیار F به عنوان معیاری جهت مقایسه عملکرد مدل‌های یاد شده، انتخاب شده‌اند. همچنین طراحی مدل ارزیابی ریسک در سطح یک سازمان سیاستگذار مانند بانک مرکزی تاکنون مورد توجه پژوهشگران داخلی و خارجی نبوده، که به دلیل ماهیت خاص نظام بانکی ایران و به علت وجود محدودیتهای ارزی ضروری است پژوهشی به صورت جامع به طراحی مدل هوشمند عمیق برای ارزیابی ریسک واردکنندگان کالا در سطح کلان بانک مرکزی بپردازد.

جدول ۱. پژوهش‌های حوزه اعتبارسنجی وارد کنندگان

ردیف	نام محقق	روش مورد استفاده	داده مورد استفاده
۱	امل و همکاران	DEA	داده‌های مرتبط با نسبتهای مالی
۲	وست	مقایسه ۲ مدل پارامتریک،۳ مدل ناپارامتریک و ٥ مدل شبکه عصبی	۲ مجموعه داده اطلاعات اعتباری آلمان و استرالیا شامل اطلاعات هویتی مانند سن، جنسیت، شغل، وضعیت مسکن، ملیت، وضعیت تاهل و اطلاعات اعتباری مانند تاریخچه اعتبار، عملکرد حساب، هدف اخذ تسهیلات، مقدار تسهیلات
۳	ابطحی و همکاران	شبکه عصبی مصنوعی	مجموع تسهیلات، مجموع سپرده، دارایی‌های نقد

۱ K-nearest Neighbors
۲ Keras Sequential neural networks

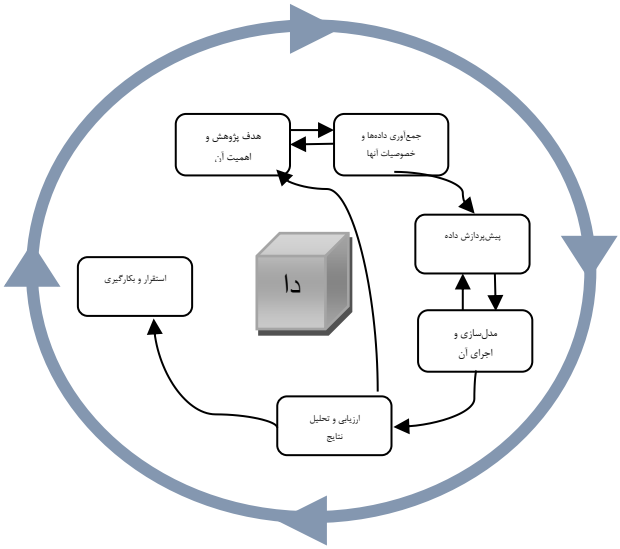
ردیف	نام محقق	روش مورد استفاده	داده مورد استفاده
٤	ونسچی و همکاران	برنامه‌ریزی ژنتیک و یادگیری ماشین	محل فعالیت مشتری، عملکرد مالی، میانگین تسهیلات
٥	اورسکی و همکاران	الگوریتم ژنتیک	داده‌های هویتی (سن، جنسیت، شغل)، داده‌های مالی (تعداد و مجموع گردش مالی، میانگین گردش حساب، تعداد و میانگین چک‌های برگشتی)، داده‌های تسهیلاتی (تاریخچه تسهیلات دریافتی، تعداد و میانگین تسهیلات دریافتی
٦	تای و همکاران	بکارگیری شبکه عصبی متوالی عمیق و شبکه عصبی پیچشی	مجموعه داده اعتباری آلمان شامل ١٠٠٠ رکورد و ٢٤ متغیر، مجموعه داده اعتباری استرالیا شامل ٦٩٠ رکورد و ١٤ متغیر و مجموعه داده اعتباری کگل شامل ٢٥٠٠٠٠ رکورد و ١٠ متغیر
٧	منگ و همکاران	جنگل تصادفی	١٤ متغیر داده‌های مالی برای بازه زمانی ٣ ساله
٨	ژائو و همکاران	جنگل تصادفی	٧ متغیر شامل میانگین درآمد سالانه، میانگین فروش سالانه، میانگین نرخ رشد سالانه سود، میانگین نرخ رشد سالانه درآمد، نرخ تبادلات غیر معتبر، بالاترین حد نوسانات، نوسانات درآمد و رتبه اعتباری
٩	وانگ و همکاران	XGBoosT	٢٦ ویژگی شامل داده‌های هویتی (سن، جنسیت، وضعیت تاهل) و سوابق مالی (سوابق تسهیلات، وضعیت درآمد، وضعیت بدهی خانوار)
١٠	لی و همکاران	شبکه عصبی عمیق	سوابق داده‌های مالی و سوابق تسهیلاتی مشتریان
١١	ژو و همکاران	یادگیری عمیق هیبریدی	سوابق مالی و تسهیلاتی
١٢	آدو و همکاران	یادگیری عمیق	داده‌های مرتبط با نسبتهای مالی و ترازنامه شرکت‌های تحت بررسی
١٣	تیاگی و همکاران	٤ مدل یادگیری ماشین (انتخاب مدل Kears بعنوان مدل منتخب)	٣٠ ویژگی مرتب با سوابق مالی برای یکصد و ده هزار رکورد
١٤	باستوس	مقایسه مدل درخت تصمیم تقویت شده با مدل شبکه عصبی چند لایه و ماشین بردار پشتیبان	مجموعه داده اعتباری آلمان شامل ١٠٠٠ رکورد و ٢٤ متغیر، مجموعه داده اعتباری استرالیا شامل ٦٩٠ رکورد
١٥	مستیری و همکارن	٦ مدل یادگیری ماشین	١٢ ویژگی از جمله سن، درآمد، وضعیت تاهل، تحصیلات، گردش حساب، میزان تسهیلات دریافتی برای ٦٨٨ فرد حقیقی
١٦	صفری و همکاران	تحلیل پوششی داده‌ها	فروش، تسهیلات بانکی، نقدینگی، فعالیت، سودآوری و بدهی

ردیف	نام محقق	روش مورد استفاده	داده مورد استفاده
۱۷	صفایی و همکاران	تاپسیس	جنسیت، تحصیلات، محل سکونت، شغل، سابق اعتباری، وضعیت چک، نرخ سود و وثیقه
۱۸	حری و همکاران	مدل ترکیبی عصبی فازی	سن، سابقه ارتباط با بانک و میزان اعتبار
۱۹	سهرابی و همکاران	مدل‌های طبقه بندی و خوشه‌بندی	از داده‌های حساب مشتری و داده‌های تسهیلاتی
۲۰	البرزی و همکاران	مدل‌های مختلف یادگیری ماشین	ویژگی‌هایی همچون سن، مبلغ و نوع وثیقه، مبلغ و تعداد تسهیلات، بخش اقتصادی
۲۱	جدوال و همکاران	مدل هیبریدی خوشه‌بندی و طبقه‌بندی	سوابق هویتی (محل فعالیت، سابقه فعالیت شرکت) و سوابق مالی (مانده حساب، گردش مالی حساب) و سوابق تسهیلاتی (تعداد و مجموع تسهیلات دریافتی)

۳- روش‌شناسی پژوهش:

همانطور که بیان شد، در ادبیات موضوع روش‌های مختلفی جهت ارزیابی ریسک توسعه داده شده‌اند. روش‌های یاد شده شامل طیفی از مدل‌های آماری و مدل‌های مبتنی بر یادگیری ماشین بوده، از جمله مهمترین مدل‌های آماری می‌توان به آنالیز تشخیص خطی، رگرسیون لجستیک و بیز ساده اشاره نمود. با این حال، زمانی که در پژوهش با داده‌های بزرگ[۱] سروکار داریم، مدل‌های آماری توانایی خود را از دست می‌دهند [۳۰]. لذا در این پژوهش از متدولوژی CRISP-DM[۲] استفاده می‌شود. کریسپ فرآیندی است که بر تشریح راهبردهای معمول داده‌کاوان خبره به منظور غلبه بر مشکلات داده‌کاوی تمرکز دارد [۳۱]. این فرآیند از ٦ مرحله اصلی تشکیل شده است. این مراحل عبارتند از: هدف پژوهش و اهمیت آن[۳]، جمع‌آوری داده‌ها و خصوصیات آنها[٤]، پیش‌پردازش داده‌ها[٥]، مدل‌سازی و اجرای آن[٦]، ارزیابی و تحلیل نتایج[۷]، استقرار و بکارگیری[۸]. توالی مراحل متدولوژی CRISP-DM همیشگی نیست وبراساس نوع پژوهش و خروجی‌های هرمرحله، قابل تغییر است [۳۱]. مراحل این مدل در نمودار۱ ارائه شده است.

[۱] Big Data
[۲] Cross-industry standard process for data mining as CRISP-DM
[۳] Business Understanding
[٤] Data Understanding
[٥] DataPreparation
[٦] Modeling
[۷] Evaluation
[۸] Deployment



نمودار ۱ : مدل فرآیندی CRISP-DM

هدف پژوهش و اهمیت آن : مرحله اول از مدل CRISP-DM شناخت کسب و کار است. این مرحله روی اهداف و الزامات پروژه متمرکز می شود و در این مرحله سعی بر آن است که مساله اصلی شناسایی گردد [۳۱].

جمع‌آوری داده‌ها و خصوصیات آنها : مرحله درک (شناخت) داده شامل نگاه نزدیک‌تر به در دسترس بودن داده برای داده‌کاوی بوده و شامل گردآوری داده‌های اولیه، توصیف داده، کشف داده، و تغییر کیفیت داده می‌شود [۳۱].

پیش‌پردازش داده‌ها : مرحله پیش‌پردازش داده‌ها در برگیرنده کلیه فعالیت‌هایی است که برای ساخت مجموعه داده نهایی از مجموعه داده‌های خام اولیه مورد نیاز بوده، به عبارتی داده‌ها پس از آماده‌سازی در این مرحله، به عنوان ورودی مرحله مدلسازی مورد استفاده قرار می‌گیرند [۳۱].

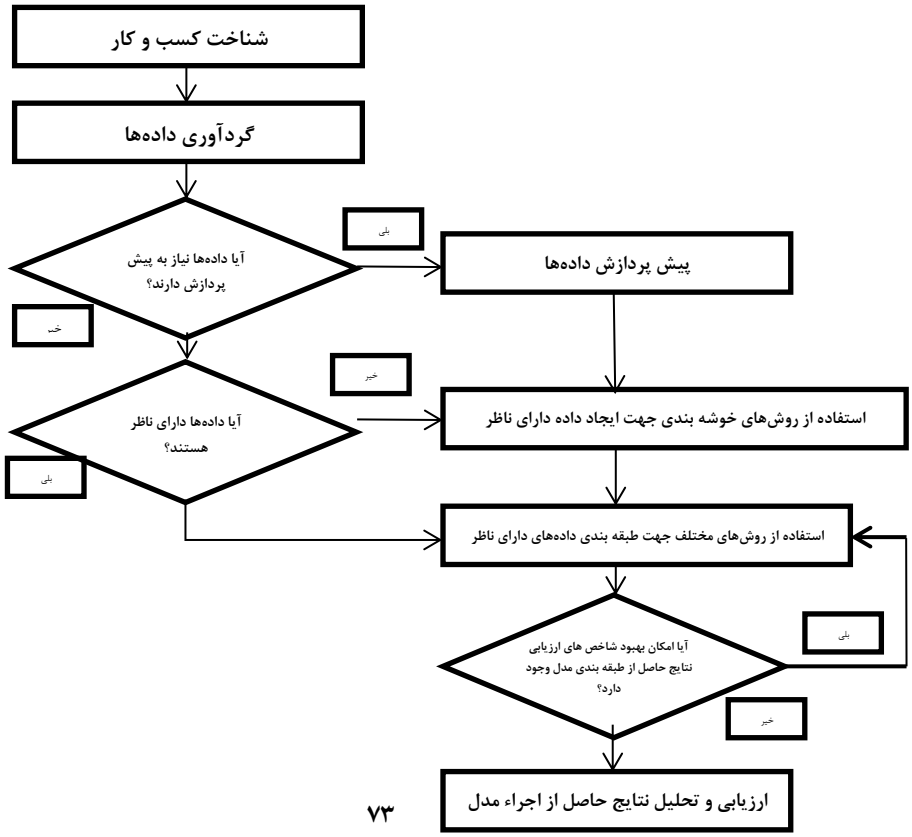
مدلسازی و اجرا: در این مرحله برای آموزش مدل، از داده‌های پردازش شده استفاده می‌شود (برزوزوفسکا و همکاران، ۲۰۲۳). با توجه به این امر که داده‌های مورد استفاده داده‌هایی بدون ناظر هستند، در ابتدا با استفاده از روش خوشه‌بندی کی‌مینز[۱] نسبت به ایجاد ناظر برای هر

[۱] K Means

داده اقدام می‌شود (عملیات لیبل‌زنی) سپس به منظور بهبود مدل مورد استفاده، داده‌های دارای ناظر (لیبل) به عنوان ورودی مدل‌های طبقه‌بندی (شامل روش‌های جنگل تصادفی، اکس جی بوست و شبکه عصبی کراس ترتیبی) مورد استفاده قرار خواهند گرفت.

ارزیابی و تحلیل نتایج : در مرحله ارزیابی، نتایج حاصل از بکارگیری هر یک از مدل‌های طبقه‌بندی مورد ارزیابی قرار می‌گیرند [۳۱].

استقرار و بکارگیری : منظور از استقرار، پیاده سازی نهایی برنامه طراحی شده به گونه‌ای است که مشتری بدون نیاز به متولی پژوهش قادر به استفاده از آن باشد. به عبارتی دانش ضمنی کسب شده از مراحل قبلی باید به نحوی مورد استفاده قرار گیرد که نتایج آن برای مدیران سازمان ملموس باشد [۳۱]. در این مرحله پس از بررسی نتایج حاصل از بکارگیری مدل و با کسب نظر از خبرگان سازمان، نسبت به تحلیل نتایج بدست آمده اقدام خواهد شد. با توجه به مراحل مدل CRISP-DM طرح کلی پژوهش به شرح نمودار۲ است.



نمودار۲. مراحل اجرای فرایند پژوهش حاضر

۴- پیاده‌سازی

همانطور که اشاره شد، مدیریت ریسک واردکنندگان کالا و اخذ وثیقه ارائه پروانه الکترونیکی گمرکی متناسب با ریسک هر واردکننده، پوشش ریسک ناشی از نوع روش پرداخت حواله ارزی، جلوگیری از تخصیص و فروش ارز به اخلال‌گران ارزی و در نتیجه مدیریت منابع ارزی به عنوان مهمترین اهداف این پژوهش بیان شده‌اند که در ادامه بدان پرداخته خواهد شد.

۴-۱- هدف پژوهش و اهمیت آن

این مرحله روی اهداف و الزامات پروژه تمرکز نموده و سعی برآن دارد که مساله اصلی شناسایی گردد. مساله اصلی این پژوهش طراحی مدل ارزیابی ریسک واردکنندگان کالا به نحوی است که مدل نهایی توانایی ارزیابی ریسک واردکنندگان و اخذ وثایق متناسب با ریسک هر واردکننده و جلوگیری از تخصیص و فروش ارز به اخلال‌گران ارزی را در اولویت دهد.

۴-۲- جمع‌آوری داده‌ها و خصوصیات آنها

داده‌های این پژوهش شامل عملکرد ارزی و ریالی واردکنندگان کالا است. بر اساس پژوهش‌های مشابه، مهم‌ترین داده‌ها برای رتبه‌بندی مشتریان عبارتند از: داده‌های هویتی (محل و سابقه فعالیت)، سوابق عملکرد حساب ریالی (تعداد، میانگین، حجم و وضعیت چک‌ها) و سوابق تسهیلاتی (میزان و تعداد). برای سوابق ارزی نیز، حجم و تعداد فعالیت‌ها و وضعیت تعهدات ارزی کلیدی‌ترین داده‌ها هستند. این داده‌ها از سامانه‌های ارزی و ریالی بانک مرکزی (به‌ویژه پورتال ارزی) استخراج و با استفاده از شناسه منحصربه‌فرد مشتری تجمیع شده‌اند، به‌گونه‌ای که هر سطر نماینده یک واردکننده و هر ستون نماینده یک ویژگی است. مجموعه داده شامل ۱۰۰,۰۰۰ رکورد و ۲۰ ویژگی است که تعریف دقیق آن‌ها در جدول ۲ آمده است.

جدول ۲. ویژگی‌های مجموعه داده‌ها

ردیف	Attribute Name	نام ویژگی	توضیحات
۱	CUSTOMER_ID	شناسه مشتری	بیانگر شماره منحصر بفرد برای هر مشتری است
۲	CUSTOMER_PROVINCE	استان مشتری	استان محل فعالیت مشتری را مشخص می‌نماید
۳	CUSTOMER_TYPE	نوع مالکیت مشتری	نوع مالکیت مشتری (حقیقی/حقوقی) را مشخص می‌کند

ردیف	Attribute Name	نام ویژگی	توضیحات
٤	CUSTOMER_SUBTYPE	نوع فعالیت مشتری	نوع فعالیت مشتری (تولیدی/تجاری) را مشخص می‌کند
٥	OPENING_DATE	تاریخ افتتاح حساب	تاریخ افتتاح حساب مشتری در شبکه بانکی را مشخص می‌کند
٦	TURNOVER_SUM	مجموع گردش مالی	مجموع گردش مالی در شبکه بانکی را مشخص می‌کند
۷	TURNOVER_COUNT	تعداد گردش مالی	تعداد گردش مالی در شبکه بانکی را مشخص می‌کند
۸	TURNOVER_AVERAGE	میانگین گردش مالی	میانگین گردش مالی در شبکه بانکی را مشخص می‌کند
۹	BOUNCED_CHECK_COUNT	تعداد چک برگشتی	تعداد چک برگشتی مشتری (رفع سوء اثر شده) در شبکه بانکی را مشخص می‌کند
۱۰	BOUNCED_CHECK_SUM	مجموع چک برگشتی	مجموع ارقام چک برگشتی مشتری (رفع سوء اثر شده) در شبکه بانکی را مشخص می‌کند
۱۱	LOAN_SUM	مجموع تسهیلات دریافتی	مجموع مانده مبالغ تسهیلات دریافتی مشتری در شبکه بانکی را مشخص می‌کند
۱۲	LOAN_COUNT	تعداد تسهیلات دریافتی	تعداد تسهیلات دریافتی مشتری در شبکه بانکی را مشخص می‌کند
۱۳	FOREIGN_CURRENCY_SUM	مجموع عملیات ارزی	مجموع عملیات ارزی مشتری در شبکه بانکی را مشخص می‌کند
۱٤	FOREIGN_CURRENCY_COUNT	تعداد عملیات ارزی	تعداد پرونده‌های ارزی مشتری در شبکه بانکی را مشخص می‌کند
۱٥	CURRENCY_COMMITMENT_SUM	مجموع تعهدات ارزی	مجموع تعهدات ارزی مشتری (منقضی شده و نشده) در شبکه بانکی را مشخص می‌کند
۱٦	CURRENCY_COMMITMENT_COUNT	تعداد تعهدات ارزی	تعداد پرونده‌های دارای تعهدات ارزی مشتری در شبکه

ردیف	Attribute Name	نام ویژگی	توضیحات
			بانکی(منقضی شده و نشده) را مشخص می‌کند
۱۷	EXPIRE_CURRENCY_COMITMNT_SUM	مجموع تعهدات ارزی منقضی شده	مجموع تعهدات ارزی منقضی شده مشتری در شبکه بانکی را مشخص می‌کند
۱۸	EXPIRE_CURRENCY_COMITMNT_COUNT	تعداد تعهدات ارزی منقضی شده	تعداد پرونده‌های دارای تعهدات ارزی منقضی شده مشتری در شبکه بانکی را مشخص می‌کند
۱۹	CURRENCY_OBLIG_RESOLVED_SUM	مجموع تعهدات ارزی منقضی نشده	مجموع تعهدات ارزی منقضی نشده مشتری در شبکه بانکی را مشخص می‌کند
۲۰	CURRENCY_OBLIG_RESOLVED_COUNT	تعداد تعهدات ارزی منقضی نشده	تعداد پرونده‌های دارای تعهدات ارزی منقضی نشده مشتری در شبکه بانکی را مشخص می‌کند

با توجه به نظر خبرگان، "مجموع تعهدات ارزی منقضی شده" مهمترین ویژگی در بین ۲۰ ویژگی تحت بررسی بوده، با این حال به منظور تعیین متغیر وابسته و با توجه به ویژگی‌های مورد نظر، عملیات لیبل‌زنی در ادامه با استفاده از روش خوشه‌بندی کی‌مینز انجام شده است. شایان ذکر اینکه تغییر در ویژگی‌ها می‌تواند منجر به تغییر در نتیجه عملیات خوشه‌بندی و طبقه‌بندی شود ولیکن در این پژوهش بهبود در میزان معیارهای صحت و دقت در اثر بکارگیری مدل‌هایِ مختلف طبقه‌بندی مدنظر می‌باشد.

همچنین بررسی اولیه مجموعه داده‌های اخذ شده بیانگر آن است که "نوع فعالیت" حدود ۱۰ درصد از مشتریان مجموعه داده‌های تحت بررسی، "تولیدی" بوده و "نوع فعالیت" سایر مشتریان (به میزان ۹۰ درصد) "تجاری" بوده است. همچنین "نوع مالکیت" حدود ٤۹ درصد از مشتریان مجموعه داده‌های تحت بررسی، "حقیقی" بوده و "نوع مالکیت" سایر مشتریان (به میزان ۵۱ درصد) "حقوقی" بوده است. همچنین بررسی "نوع فعالیت" مشتریان "حقیقی" بیانگر این امر بوده که این دسته از مشتریان صرفاً دارای نوع فعالیت "تجاری" بوده‌اند. انتخاب مجموعه داده‌ها به نحوی بوده است که مشتریان از کلیه استان‌های کشور انتخاب شده‌اند. هر

چند که فراوانی مشتریان در استان‌های مختلف متفاوت بوده و استان تهران بیشترین فراوانی مشتریان ارزی و استان سیستان و بلوچستان کمترین فراوانی مشتریان ارزی را داشته‌اند.

۴-۳- پیش پردازش داده‌ها

کیفیت داده‌ها یک عامل مهم در استخراج بینش‌های معنادار و تصمیم‌گیری آگاهانه است. پیش پردازش داده[۱] همچنین به‌عنوان پاکسازی داده[۲]نیز شناخته می‌شود. این عملیات نقشی حیاتی در تبدیل داده‌های خام به قالبی تمیز و ساختاریافته ایفا می‌کند. همانطور که پیش از این بیان شد مجموعه داده‌های این پژوهش از سامانه‌های ارزی و ریالی بانک مرکزی اخذ شده است. بنابراین پس از واکشی داده‌های مورد نیاز از سامانه‌های ارزی و ریالی از جمله پورتال ارزی، پیش‌پردازش داده‌ها (شامل جایگزینی مقادیر از دست رفته، تبدیل داده‌ها و باز نمونه‌گیری) در صورت نیاز اجرا می‌گردد. در گام نخست مشخص گردید مجموعه داده‌های مورد بررسی فاقد مقادیر از دست رفته بوده، ولیکن در بررسی داده‌های ارزی مشخص گردید که برخی از داده-های مرتبط با ویژگی "مجموع عملیات ارزی"، دارای مقادیر بسیار بالا می‌باشند. به منظور جلوگیری از تاثیر داده‌های مذکور در فرآیند تجزیه و تحلیل، داده‌های یاد شده با میانگین مجموعه داده‌ها جایگزین شده‌اند. با توجه به اینکه بخشی از داده‌ها از سامانه‌های ریالی و بخشی از داده‌ها از سامانه‌های ارزی اخذ شده‌اند، داده‌ها با توجه به ویژگی‌هایی مانند شماره اقتصادی و شماره کارت بازرگانی، یکپارچه شده و عمل "یکپارچه سازی داده‌ها" انجام شده و در نتیجه هر سطر بیانگر ویژگی های رفتاری و عملکرد ارزی و ریالی مشتری است. در نهایت باید به این نکته توجه نمود که در مجموعه داده‌های تحت بررسی از داده‌های ریالی و ارزی استفاده شده و با توجه به تفاوت ارزش این دو نوع داده، بی‌مقیاس‌سازی داده‌ها ضروری است. لذا به منظور حذف تاثیر تفاوت ارزش مقادیر ریالی و ارزی داده‌های مجموعه با استفاده از تابع Standard.Scaler در پایتون بی‌مقیاس شده‌اند.

۴-۴- مدل‌سازی و اجرا

این مرحله پس از گردآوری مجموعه داده‌ها و پس از مرحله پیش پردازش داده‌ها صورت می‌پذیرد. نگاهی دقیق‌تر به مجموعه داده‌ها، این امر را خاطر نشان می‌سازد که مجموعه داده-های یاد شده، شامل داده‌های بدون ناظر است. همانطور که پیش از این بیان شد، تحلیل داده‌های

---

[۱] Data Preprocessing [۲] Data wrangling

بدون ناظر با استفاده از روش‌های خوشه‌بندی صورت می‌پذیرد. با این حال روش‌های خوشه‌بندی امکان بهبود را ندارند. به منظور بهبود در میزان دقت و صحت، در ابتدا داده‌های بدون ناظر با استفاده از روش‌های خوشه‌بندی به داده‌های با ناظر تبدیل شده، سپس با بکارگیری روش‌های مختلف طبقه‌بندی، نسبت به بررسی میزان معیارهای ارزیابی هر یک از مدل‌ها و انتخاب مدلی با بالاترین میزان معیار ارزیابی اقدام می‌شود.

از سویی دیگر، "نوع فعالیت مشتریان" تحت بررسی (تولیدی/تجاری) می‌تواند بر فرآیند مدل‌سازی تاثیرگذار باشد. بررسی مجموعه داده‌ها بیانگر آن است که هر دو نوع مشتریان به واردات کالا اقدام می‌نمایند. با این حال مشتریان با نوع فعالیت "تجاری" به صورت مستمر به فرآیند واردات کالاهای گوناگون اقدام نموده، در صورتیکه کالاهای وارداتی مشتریان با نوع فعالیت "تولیدی" اغلب شامل خطوط تولید و مواد اولیه تولیدی است. همچنین همانطور که پیش‌تر بیان شد، نوع فعالیت حدود ۱۰ درصد از مشتریان تحت بررسی "تولیدی" بوده و حدود ۹۰ درصد از مشتریان مجموعه داده تحت بررسی، مشتریان با نوع فعالیت "تجاری" را شامل می‌شوند. هر دو نوع مشتریان تحت بررسی موظف به ترخیص کالا در مهلت مقرر می‌باشند. با این حال مهلت مجاز جهت ترخیص کالا (از زمان تامین ارز تا زمان ترخیص کالا (تاریخ رسید مالی مندرج در پروانه الکترونیکی ترخیص کالا)) برای مشتریان با نوع فعالیت "تجاری" ۸ ماه و برای مشتریان با نوع فعالیت "تولیدی" از ۱۱ الی ۲۸ ماه بوده [۳] و این تفاوت مهلت بر میزان "مجموع تعهدات ارزی منقضی شده" هر دو گروه مشتریان تاثیرگذار خواهد بود.

لذا با توجه به نامشخص بودن تاثیر نوع فعالیت مشتری بر "طراحی مدل هوشمند اعتباریابی واردکنندگان کالا با رویکرد یادگیری ماشین"، طراحی مدل یاد شده در مرحله نخست بدون در نظر گرفتن تفاوت "نوع فعالیت مشتریان" (تولیدی/تجاری) و سپس با در نظر گرفتن تفاوت "نوع فعالیت مشتریان"صورت خواهد پذیرفت. در انتها، میزان معیارهای ارزیابی هر دو مدل مورد بررسی خبرگان قرار گرفته، و با توجه به میزان تاثیر ویژگی "نوع فعالیت مشتریان" بر عملکرد "مدل هوشمند اعتباریابی واردکنندگان کالا با رویکرد یادگیری ماشین" مدل نهایی انتخاب خواهد شد.

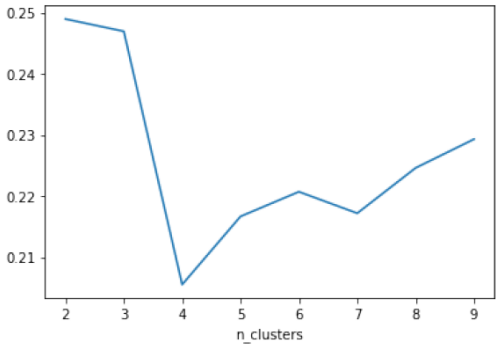
۴-۴-۱- خوشه‌بندی و ایجاد مجموعه داده با ناظر

اشاره شد که به منظور بهبود مدل، نیاز به داده‌های با ناظر است. درحالیکه مجموعه داده‌های تحت بررسی فاقد ناظر بوده که این امر امکان بررسی مدل‌های مختلف و امکان بهبود میزان صحت و دقت مدل را محدود می‌نماید. لذا به منظور رفع این مشکل، ابتدا نسبت به تبدیل مجموعه داده بدون ناظر به مجموعه داده با ناظر اقدام و بدین منظور از روش خوشه‌بندی کی‌مینز استفاده شده است. الگوریتم کی‌مینز یک الگوریتم یادگیری بدون نظارت است که برای حل مشکلات خوشه بندی در علم داده استفاده می‌شود و مبتنی بر کمی‌سازی بردارهاست. این الگوریتم قصد دارد یک مجموعه داده را به K خوشه مجزا افراز کند به نحوی که هر خوشه نشان‌دهنده گروهی از نقاط داده‌ای است که شباهت‌های مشترکی دارند و امکان بینش معنادار و کشف الگو را فراهم می‌کنند [۳۲]. با این حال باید در نظر داشت که به منظور انجام خوشه-بندی، ضروریست تعداد خوشه‌ها از قبل مشخص باشد. تعداد خوشه‌ها در عمل خوشه‌بندی را می‌توان با استفاده از نظرات خبرگان و یا با استفاده از روش‌های علمی تعیین نمود. در این پژوهش به منظور تعیین تعداد خوشه‌ها از روش‌های علمی استفاده شده است. روش‌های ریاضی مورد استفاده جهت تعیین تعداد خوشه‌ها شامل روش سیلوهت[۱][۳۳] ، روش کالینسکی-هاراباسز[۲][۳٤] و روش دیویس-بالدین[۳][۳۵] می‌باشند. در پژوهش حاضر به منظور بکارگیری روش‌های تعیین تعداد خوشه و اجرای خوشه‌بندی از روش خوشه‌بندی کی‌مینز در نرم‌افزار پایتون استفاده شده است. نتایج بکارگیری ۳ مدل یاد شده به شرح زیر است :

الف) در صورت تفکیک مجموعه داده‌ها براساس ویژگی "نوع فعالیت مشتریان" به ۲ مجموعه داده "تجاری" و "تولیدی"، روش سیلوهت و روش دیویس-بالدین ۲ خوشه و روش کالینسکی-هاراباسز ۳ خوشه را برای مشتریان با نوع فعالیت "**تجاری**" در نظر گرفته‌اند که با توجه به اینکه در روش‌های مدیریت ریسک حداقل ۳ خوشه مورد استفاده قرار می‌گیرد، در نهایت ۳ خوشه برای خوشه‌بندی در نظر گرفته شده است. نتایج بکارگیری ۳ روش یاد شده در نمودار ۳ مشخص شده است.

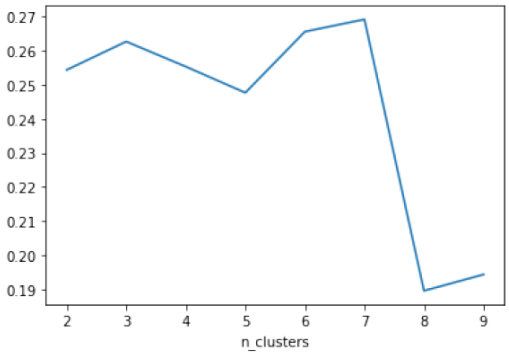
---

[۱] Silhouette [۲] Callinski-Harabasz [۳] Davies-Bouldin



نمودار ۳. تعداد خوشه برای مشتریان با نوع فعالیت "تجاری"

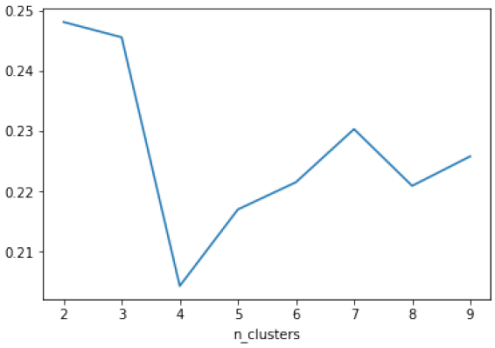
ب) در صورت تفکیک مجموعه داده‌ها براساس ویژگی "نوع فعالیت مشتریان" به ۲ مجموعه داده "تجاری" و "تولیدی"، روش سیلوهت ۷ خوشه، روش کالینسکی-هاراباسز ۲ خوشه و روش دیویس-بالدین ۲ خوشه را برای مشتریان با نوع فعالیت "**تولیدی**" در نظر گرفته‌اند که در نهایت ۷ خوشه برای خوشه‌بندی در نظر گرفته شده است. نتایج بکارگیری ۳ روش یاد شده در نمودار ٤ مشخص شده است.



نمودار ۴. تعداد خوشه برای مشتریان با نوع فعالیت "تولیدی"

ج) در صورت عدم تفکیک مجموعه داده‌ها براساس ویژگی "نوع فعالیت مشتریان" به ۲ مجموعه داده "تجاری" و "تولیدی" و به عبارتی در نظر گرفتن کل مجموعه داده‌ها، روش سیلوهت ۱ خوشه، روش روش کالینسکی-هاراباسز ۲ خوشه و روش دیویس-بالدین ۱ خوشه را برای کل مجموعه داده‌های مشتریان در نظر گرفته‌اند که با توجه به اینکه در روش‌های

مدیریت ریسک حداقل ۳ خوشه مورد استفاده قرار می‌گیرد، در نهایت ۳ خوشه برای خوشه-بندی در نظر گرفته شده است. نتایج بکارگیری ۳ روش یاد شده در نمودار ۵ مشخص شده است.



نمودار۵. تعداد خوشه برای کل مشتریان بدون در نظر گرفتن ویژگی "نوع فعالیت مشتری"

۴-۴-۲- طبقه بندی و بهبود عملکرد مدل

در این مرحله پس از انجام فرآیند خوشه‌بندی و تبدیل داده‌های بدون ناظر به داده‌های با ناظر، خروجی مدل خوشه‌بندی کی‌مینز به عنوان ورودی مدل‌های مختلف طبقه‌بندی مورد استفاده قرار گرفته و پس از بکارگیری ۳ مدل مختلف طبقه‌بندی (شامل روش‌های جنگل تصادفی، اکس جی بوست و شبکه عصبی کراس ترتیبی) مدلی که میزان معیار F آن از بقیه مدل‌ها بیشتر باشد به عنوان مدل نهایی مرحله طبقه بندی مورد انتخاب قرار خواهد گرفت. همچنین باید در نظر گرفت که در مرحله طبقه‌بندی، مجموعه داده‌های تحت بررسی به ۳ قسمت شامل مجموعه داده‌های آموزش[۱]، مجموعه داده‌های توسعه[۲] و مجموعه داده‌های آزمایش[۳] تقسیم می‌شوند. در این پژوهش ، ۰.۱ مجموعه داده‌های تحت بررسی به منظور آزمایش در نظر گرفته شده‌اند و همچنین از ۰.۹ باقی مانده مجموعه داده، ۰.۷ برای آموزش و ۰.۳ برای توسعه مد نظر قرار گرفته‌اند (پیوست ۲). پس از مشخص شدن نحوه تقسیم‌بندی مجموعه داده‌ها، پارامترهای مدل‌های طبقه‌بندی مشخص می‌گردند.

[۱] Train [۲] Develop [۳] Test

در ابتدا پارامترهای مدل اکس جی بوست معرفی می‌گردند. این الگوریتم بر پایه درخت‌های تصمیم بنا شده و از تکنیک تقویت گرادیان برای دستیابی به دقت و کارایی بالاتر بهره می‌برد. پارامترهای مورد استفاده مدل اکس جی بوست در این پژوهش به شرح زیر عبارتند از :

booster = 'gbtree' , پارامتر تقویت کننده

n_estimators = ۵ , تعداد فراگیران موجود در درخت تصمیم گیری

num_parallel_tree = ۲۰۰ , تعداد درخت‌های موازی ساخته شده در هر تکرار

verbosity = ۰ سطح تکرار چاپ پیام‌ها

حال به بررسی پارامترهای مدل جنگل تصادفی می‌پردازیم. جنگل تصادفی نوعی الگوریتم یادگیری گروهی[۱]است که چندین درخت تصمیم را برای پیش‌بینی ترکیب می‌کند. پارامترهای مورد استفاده مدل جنگل تصادفی در این پژوهش به شرح زیر عبارتند از :

n_estimators = ۲۰۰ , تعداد فراگیران موجود در درخت تصمیم گیری

max_depth = ۸, حداکثر عمق درخت که افزایش آن می‌توان منجر به پیچیده‌تر شدن مدل شود

verbosity = ۰ سطح تکرار چاپ پیام‌ها

در نهایت پارامترهای مدل طبقه‌بندی کراس-ترتیبی که در این پژوهش مورد استفاده قرار گرفته‌اند، معرفی می‌گردند. مدل کراس-ترتیبی یک کتابخانه یادگیری عمیق منبع باز محسوب می‌شود. این مدل آزمایش سریع با استفاده از مدل‌های یادگیری عمیق را امکان‌پذیر می‌کند. پارامترهای مدل کراس-ترتیبی در این پژوهش به شرح زیر عبارتند از :

```
model۱ = tf.keras.Sequential()
model۱.add(tf.keras.layers.Input(shape = (۱۹,)))
model۱.add(tf.keras.layers.Dense(۵۰ , activation = "relu"))
model۱.add(tf.keras.layers.Dense(۱۰۰ , activation = "relu"))
model۱.add(tf.keras.layers.Dense(۷ , activation = 'softmax'))
return model۱
```

اولین پارامتر پس از فراخوانی کراس-ترتیبی، تعداد گره‌ها است که در این پژوهش تعداد گره در لایه اول ۵۰ در نظر گرفته شده است. پارامتر بعدی تابع "فعال‌سازی"[۲]است. در اینجا از تابع "یکسوساز"[۳] که به اختصار "relu" نامیده شده است به عنوان تابع فعال‌سازی استفاده می‌شود. همچنین پیکربندی فرآیند یادگیری در این پژوهش به شرح زیر است :

```
Model_tolidi.compile(optimizer='adam',
loss='categorical_crossentropy',metrics=['accuracy'])
```

---

[۱] Ensemble Learning Algorithm

[۲] Activation Function

[۳] Rectifier

به منظور پیکربندی مدل کراس–ترتیبی در این پژوهش از بهینه ساز[۱] آدام[۲] استفاده شده است. پس از پیکربندی مدل، می‌توان مراحل آموزش را شروع نمود. این عمل در مدل کراس–ترتیبی با استفاده از تابع model.fit() صورت می‌پذیرد. در این مرحله باید داده های آموزش، اندازه دسته و تعداد دوره ها را مشخص کرد. مدل کراس–ترتیبی به طور خودکار نحوه انتقال داده ها را به صورت تکراری به بهینه ساز برای تنظیم تعداد دوره ها، مشخص می کند. اندازه دسته در این پژوهش ۱۰۰ و تعداد دوره‌های تکرار ۲۰ انتخاب شده است.

```
Model_tolidi.fit(data_tolidi_train , label_tolidi_train_encode , epochs = ۲۰ ,
batch_size = ۱۰۰)
Model_tejari.fit(data_tejari_train , label_tejari_train_encode , epochs = ۲۰ ,
batch_size = ۱۰۰)
Model_kol.fit(data_kol_train , label_kol_train_encode , epochs = ۲۰ ,batch_size =
۱۰۰)
```

۴-۵- ارزیابی و تحلیل نتایج

در این مرحله پس از اجرای مدلهای طبقه‌بندی و به منظور تحلیل نتایج از معیار F استفاده شده است. معیار F به صورت زیر تعریف می‌گردد.

(فرمول شماره ۱)

$$F = \frac{۲*\text{Sensitivity}*\text{Precision}}{\text{Sensitivity}+\text{Precision}}$$

در فرمول فوق، معیار حساسیت[۳] بیانگر نسبت نمونه‌های مثبتی است که به درستی پیش‌بینی شده‌اند (TP) به کل نمونه‌های مثبت، که به صورت زیر تعریف می‌گردد :

(فرمول شماره ۲)

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN}$$

همچنین معیار دقت[٤] بیانگر نسبت نمونه‌هایی است که به درستی در طبقه مثبت قرارگرفته‌اند. فرمول معیار دقت عبارت است از:

(فرمول شماره ۳)

$$\text{Precision} = \frac{TP}{TP+FP}$$

مقادیر فرمول‌های فوق به شرح تعریف می‌گردند :

TP[۵]: تعداد نمونه‌های متعلق به طبقه اقلیت (طبقه مثبت) که مدل به درستی آنها را پیش‌بینی کرده است.

---

۱ Optimizer
۲ Adam
۳ Sensitivity
٤ Precision
۵ True positive

TN[1]: تعداد نمونه‌های متعلق به طبقه اکثریت (طبقه منفی) که مدل به درستی آنها را پیش‌بینی کرده است.

FP[2] : تعداد نمونه‌های متعلق به طبقه منفی که مدل به اشتباه پیش‌بینی کرده است.

FN[3] : تعداد نمونه‌های متعلق به طبقه مثبت که مدل به اشتباه پیش‌بینی کرده است.

حال با توجه به تعریف معیار F مدل‌های مختلف طبقه‌بندی با توجه به معیار یاد شده مورد بررسی و مقایسه قرار خواهند گرفت. میزان معیار F پس از بکارگیری مدل جنگل تصادفی به شرح ستون اول جدول شماره ۳ بوده و همانطور که مشخص است میزان این معیار برای زمانی که مشتریان با "نوع فعالیت" تجاری و تولیدی از یکدیگر تفکیک می‌گردند تفاوت معنی‌داری نداشته و اختلاف مقادیر F بسیار ناچیز است.

جدول ۳. نتایج معیار F برای مدل‌های مختلف طبقه بندی

	RANDOM FOREST	XGBOOST	keras.Sequential
F1-score tolidi train :	۰,۹۷۷۳٤۰٤۷۰۷۱۹٦۱۱۸	۰,۹۹۰۸۳۰۲۱٥۱٥۲٤۹۳٦	۰,۹۹٤۳٦۱٤۸٦٦۸۸٤۲۸۲
F1-score tolididev :	۰,۹٥۳۰٤۷۲۲۱۲۷۳۸۰٦۷	۰,۹٦۲۷٥۳۳٦٥٥۱۳٦٥٦۱	۰,۹۸۲۳٦٤٦۲۳۲۷۰٤٤٤۲
F1-score tolidi test :	۰,۹٤٤۸۳۱۷۹۳۷٤۸۹۹۸٤	۰,۹٦۰۳۳۰٥۲۲٥٦٤۹۲۱۰	۰,۹۹۰۲٤۱۹۳۱٦٥۱٤۷۳۹
F1-score tejari train :	۰,۹۸٦۹۳٤٤۰۹۸٤۸٤۹۲۹	۰,۹۸۷۹۰۳٤۱٥٦۸۷۹٤٦٦	۰,۹۹۸۰۱۹۸۸۷۳٤۳۹۸۲٤
F1-score tejaridev :	۰,۹۷۹۷٦٥٥٦۲۰۱۱۹٤۰۸	۰,۹۸۲۱٥٥۸۳۳۹۸۸۷٤٤۱	۰,۹۹٦۹۸۹٥۸٤۳۰۸۳۱٦۷
F1-score tejari test :	۰,۹۷۹۷۲۷٤۳۲۷۳۳٦۹۱۱	۰,۹۸۱۳۷۱۳۳۱٦٦۲۸٥۹۷	۰,۹۹۷۸۸۳۰۲۲٦٤۰٥٤۲۷
F1-score kol train :	۰,۹۹٦٤٤۲۰٥٦٤۹۳۹٤۷۳	۰,۹۸۷۸۰٤٤۲۰٤۷٦۸۷۹۸	۰,۹۹۷۸٦۰٥٦۳۹۷۳٥۳٥٤
F1-score koldev :	۰,۹۸۸۱۳٥۰۳۳٤٤۷۳۳۹۳	۰,۹۸۳۸۸۱٥۰۸٥۱٤۹٦۹۸	۰,۹۹٦۸۹۳۹٦۹۱۱٥۹۱۰۰
F1-score kol test :	۰,۹۸۹۲٦۲۱۰۰۸٦٦٤٤۹۸	۰,۹۸٥۸۷۳۲۷٥۰۷۸۷٥٤٦	۰,۹۹٦۳۰٤۹٦۹٦۹۱۷۹۲۹

همچنین میزان معیار F پس از بکارگیری مدل اکس جی بوست به شرح ستون دوم جدول شماره ۳ بوده و همانطور که مشخص است میزان این معیار برای هنگامیکه مشتریان با "نوع فعالیت" تجاری و تولیدی از یکدیگر تفکیک می‌گردند تفاوت معنی‌داری نداشته و اختلاف مقادیر F بسیار ناچیز است. با این حال، میزان مقادیر معیار F برای مدل طبقه‌بندی اکس جی بوست

[1] True negative [2] False positive [3] False negative

از مدل طبقه‌بندی جنگل تصادفی در حالتی که مشتریان با توجه به ویژگی "نوع فعالیت" از یکدیگر تفکیک می‌گردند، بالاتر است. از سویی دیگر در صورت عدم تفکیک مشتریان با توجه به ویژگی یاد شده، نتایج حاصل از نتایج مدل اکس جی بوست بالاتر است. در نهایت، میزان معیار F پس از بکارگیری مدل کراس-ترتیبی به شرح ستون سوم جدول شماره ۳ بوده و همانطور که مشخص است و مانند دو روش قبل، میزان این معیار هنگامیکه مشتریان با "نوع فعالیت" تجاری و تولیدی از یکدیگر تفکیک می‌گردند، تفاوت معنی‌داری نداشته و اختلاف مقادیر F بسیار ناچیز است. با این حال، میزان مقادیر معیار F برای مدل طبقه‌بندی کراس-ترتیبی از نتایج مدل طبقه‌بندی اکس جی بوست و مدل طبقه‌بندی جنگل تصادفی بالاتر بوده، و بیانگر بهبود نتایج مدل خواهد بود.

نتایج ارزیابی مدل‌های مختلف طبقه‌بندی بر اساس معیار F نشان داد که ویژگی «نوع فعالیت مشتری» تأثیر معنی‌داری بر طراحی مدل هوشمند اعتباریابی واردکنندگان کالا ندارد؛ بنابراین مجموعه داده‌ها بر اساس این ویژگی تفکیک نمی‌شوند و کل داده‌ها به صورت یکپارچه استفاده می‌گردد. به عبارت دیگر، تفکیک داده‌های تجاری و تولیدی و اعمال سه مدل اصلی، تغییر چشمگیری در میزان صحت و دقت ایجاد نمی‌کند. همچنین در بررسی یکپارچه داده‌ها، مدل کراس-ترتیبی بالاترین معیار F را داشته است. از آنجا که هدف افزایش و بهبود شاخص F (یعنی صحت و دقت) بود، این مدل به عنوان مدل نهایی طبقه‌بندی انتخاب می‌گردد. ویژگی‌های مدل توسعه‌یافته در جدول ٤ ارائه شده است.

جدول ۴. مشخصات مدل کراس-ترتیبی

سایر مشخصات	Activator	تعداد گره مورد استفاد	مشخصات لایه
اندازه دسته	-	۱۹	لایه ورودی
۱۰۰	Relu	۵۰	لایه میانی اول
تعداد دوره های تکرار	Relu	۱۰۰	لایه میانی دوم
۲۰	Soft max	۷	لایه آخر

با توجه به جدول فوق، مدل نهایی دارای لایه ورودی با ۱۹ گره، لایه میانی اول با ۵۰ گره، لایه میانی دوم با ۱۰۰ گره و لایه آخر شامل ۷ گره است. همچنین در لایه‌های میانی از فعال‌کننده Relu و در لایه آخر از فعال‌کننده Soft max استفاده شده است. اندازه دسته ۱۰۰ در نظر گرفته شده و مدل پس از ۲۰ تکرار به نتیجه رسیده و همگرا شده است.

۴-۶- استقرار و بکارگیری

برای استقرار و بکارگیری نتایج مدل، از نظرات خبرگان استفاده شده است؛ با توجه به انتخاب مدل کراس-ترتیبی به عنوان مدل نهایی طبقه‌بندی و تحلیل داده‌های هر طبقه، جدول ۵ توسط خبرگان به عنوان ملاک تعیین میزان وثایق نقد و غیرنقد جهت ارائه وثیقه پروانه الکترونیکی ترخیص کالا تأیید شده است. در این جدول، داده‌ها به سه کلاس A، B و C تقسیم شده‌اند: مشتریان کلاس A (بهترین عملکرد) حداکثر ۵ درصد وثیقه نقدی و حداقل ۳۰ درصد وثیقه غیرنقدی دریافت می‌کنند؛ مشتریان کلاس C (بدترین عملکرد) حداقل ۳۰ درصد وثیقه نقدی و حداکثر ۵ درصد وثیقه غیرنقدی دریافت می‌کنند تا از ورود کالا توسط ایشان اطمینان حاصل شود؛ و مشتریان کلاس B ترکیبی از وثایق نقد و غیرنقد در بازه ۵ تا ۳۰ درصدی به گونه‌ای اخذ می‌کنند که مجموع وثایق نقد و غیرنقد دریافتی ۳۵ درصد باشد.

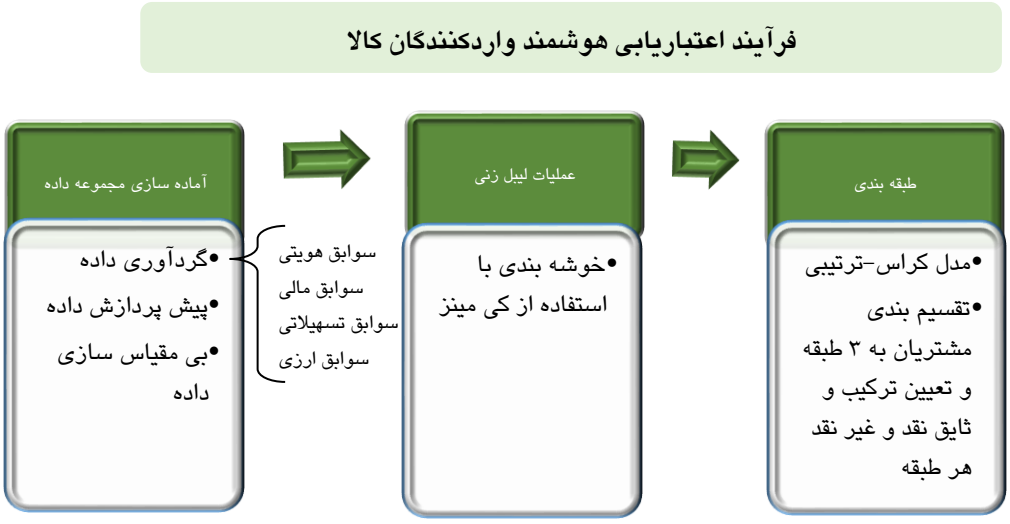
جدول۵. جدول تعیین ترکیب وثایق نقد و غیر نقد مشتریان ارزی

نحوه تعیین ترکیب وثایق نقد و غیر نقد مشتریان ارزی					
	میزان وثیقه نقد (درصد)		میزان وثیقه غیر نقد (درصد)		مجموع وثیقه (درصد)
نام کلاس	حداقل	حداکثر	حداقل	حداکثر	
کلاس A	۰	۵	۳۰	۳۵	۳۵
کلاس B	۵	۳۰	۵	۳۰	۳۵
کلاس C	۳۰	۳۵	۰	۵	۳۵

در پایان، نتایج طبقه‌بندی صورت پذیرفته توسط مدل نهایی، با استفاده از تعدادی از داده‌های عملکردی و توسط خبرگان بررسی و تحلیل شده‌اند. به عبارتی سوابق هویتی، عملکرد ریالی و ارزی ایشان مورد تجزیه و تحلیل قرار گرفته و سوابق یاد شده با طبقه تعیین شده برای ایشان مقایسه شده است. نتایج بررسی و تحلیل ذکر شده توسط خبرگان بیانگر آن است که مدل توانایی شناسایی و تفکیک مشتریان پر ریسک از مشتریان کم ریسک را داشته و در نتیجه می‌توان با اخذ وثایق نقد بیشتر، از ریسک عدم رفع تعهد ارزی مشتریان قرار گرفته در کلاس C کاسته و از سویی دیگر با اخذ وثایق غیر نقد از مشتریان طبقه A فرآیند واردات توسط ایشان را تسهیل نمود.

۵- نتیجه‌گیری

در این پژوهش، پس از بررسی پیشینه داخلی و خارجی و مطالعه مدل‌ها و ویژگی‌های مرتبط، داده‌های عملکردی مشتریان از سامانه‌های ارزی و ریالی بانک مرکزی گردآوری و پیش‌پردازش شد. به دلیل عدم وجود داده ناظر، با استفاده از روش خوشه‌بندی کی-مینز، داده‌های بدون ناظر به داده‌های ناظر تبدیل شدند تا امکان استفاده از روش‌های طبقه‌بندی فراهم گردد. سپس روش‌های مختلف طبقه‌بندی شامل جنگل تصادفی، اکس‌جی‌بوست و شبکه عصبی عمیق کراس-ترتیبی مورد ارزیابی قرار گرفتند و مدل کراس-ترتیبی به دلیل داشتن بالاترین میزان صحت و دقت (معیار F) به عنوان مدل نهایی انتخاب شد. همچنین بررسی تأثیر ویژگی «نوع فعالیت مشتری» (تولیدی/تجاری) نشان داد که تفکیک داده‌ها بر اساس این ویژگی تأثیر معنی‌داری بر شاخص F ندارد، بنابراین ضرورتی به این تفکیک وجود ندارد. نهایتاً فرآیند اعتباریابی هوشمند واردکنندگان کالا بر اساس نتایج حاصل در نمودار ٦ ارائه شده است.



نمودار ۶. فرآیند اعتبار یابی هوشمند وارد کنندگان کالا

در نهایت پیشنهاد می‌شود مدل توسعه داده شده در یکی از بانک‌های شبکه بانکی کشور به صورت نمونه مورد استفاده قرار گرفته و نتایج و تاثیر طبقه‌بندی مشتریان بر کاهش تعهدات ارزی مشتریان آن بانک (پس از یک مقطع زمانی حداقل یکساله) مورد تجزیه و تحلیل قرار

گیرد و با رفع نواقص احتمالی، نتایج مدل طبقه بندی نهایی در سامانه‌های ارزی بانک مرکزی ج.ا.ا اعمال گردد. از سویی دیگر تاثیر تغییرات مداوم و سریع قوانین و مقررات ارزی باید مد نظر قرار گرفته شود. به عنوان مثال تغییر در "حداکثر زمان مجاز به منظور ترخیص کالا"، تغییر در "درصد تعهدات ارزی منقضی شده مجاز"، تاثیر عدم رفع تعهد ارزی صادراتی مشتری بر امکان واردات کالا توسط ایشان و سایر موارد پیش‌بینی نشده بر نتایج مدل نهایی تاثیر گذار خواهد بود. لذا پیشنهاد می‌شود در زمان توسعه مدلِ نمونه، تغییرات مقررات ارزی مد نظر قرار گرفته شوند.

در صورت بکارگیری مدل نهایی و ضمن توجه به محدودیتهای فوق‌الذکر، انتظار بر آن است که زمینه فعالیت برای واردکنندگان واقعی فراهم شود و هزینه سوء استفاده از حاشیه نرخ ارز توسط اخلال‌گران ارزی افزایش یافته و زمینه مدیریت بهتر منابع ارزی فراهم شود. همچنین با توجه به تاثیر داده‌های عملکردی بر طبقه‌بندی مشتری، انتظار بر آن است که واردکنندگان کالا نسبت به رفع تعهد ارزی و بهبود عملکرد ریالی خود توجه بیشتری داشته باشند.

۶- منابع :

[۱] Asadi Zeydabadi, Fatemeh., Salmanpour, Ali., Khorshidi, Gholamhosein., Memarnejad, Abbas., Dayyani, Mehdi.(۲۰۲۳). The effects of reforming foreign trade strategies on full employment (evidence from Iran), Quarterly Journal of Islamic Economics and Banking, Vol. ٤٥, pp.٣٤٥-٣٦٣, In Persian.

[۲] Summary report on the country's foreign trade performance, Ministry of Industry, Mines and Trade, Trade Development Organization.(۲۰۲۳), In Persian.

[۳] The first set of foreign exchange regulations of the country, the Central Bank of the Islamic Republic of Iran.(۲۰۲۲), In Persian.

[٤] Bafandeh Imandoost, Sadegh., Shaterian, Zahra., Fahimi Fard, Seyyed Mohammad.(۲۰۱٦). Study of factors affecting the collection rate of facilities of the Agricultural Bank of Khorasan Razavi Province, application of the econometric model, Two Quarterly Journals of Monetary and Financial Economics, Vol. ۱۲, pp.۱۸۹-۲۱٦,In Persian.

[٥] Mojahed, Megdi.(۲۰۱٦). Typology of currency crimes as crimes against the country's economic security, Quarterly Magazine of Security Perspectives, Vol. ۳۳, pp.٥-۳۱,In Persian.

[٦] Mazini, Masoud., Mohajerani Tehrani, Mohammad Hassan.,(۲۰۲۳).International Banking ۲,Education institutionof Banking, Central Bank of the Islamic Republic of Iran, In Persian.

[۷] Faragmand Moeen, Hamed., Samavi, Mohammad Ebrahim., Kousha, Emad.(۲۰۱۸). Prioritizing Iranian International Banking Payment Methods Using Analytic Hierarchy

Process, Quarterly Journal of Islamic Finance and Banking Studies. Vol. ٤,No.١٠, pp.١١٥-١٣٤,In Persian.

[٨] Shahrami Babakan, Majid., Saranj, Alireza., Nadiri, Mohammad., Noorbakhsh, Asgar.(٢٠٢٣). Comparison of factors affecting credit risk of different groups in Iranian banking system,Journal of econometric modeling, Vol. ٨,No.٢, pp.٦٩-٩٥,In Persian.

[٩] Emel, A.B., Oral, M., Reisman,A.,Yolalan,R.(٢٠٠٣). Acredit scoring approach for the commercial banking sector, Socio-economic planning science, Vol. ٣٧, pp.١٠٣-١٢٣. doi:١٠,١٠١٦/S٠٠٣٨-٠١٢١(٠٢)٠٠٠٤٤-٧

[١٠] West, D. (٢٠٠٠). Neural network credit scoring models. Computers & Operations Research, ٢٧(١١–١٢), ١١٣١–١١٥٢. DOI:١٠,١٠١٦/S٠٣٠٥-٠٥٤٨(٩٩)٠٠١٤٩-٥

[١١] Tavana, M., Abtahi, A.R., Caprio, D.D., Poortarigh, M.(٢٠١٨). An artificial neural network and bayesian network model for liquidity risk assessment in banking, Neurocomputing Journal, Vol.٢٧٥, pp.٢٥٢٥-٢٥٥٤.https://doi.org/١٠,١٠١٦/j.neucom.٢٠١٧,١١,٠٣٤

[١٢] Vanneschi ,L. ,Horn ,D.M., Castelli, M., Popovic ,A.(٢٠١٨) An Artificial intelligence system for predicting customer default in e-commerce, Expert systems with application Journal, Vol.١٠٤, pp.١-٢١.https://doi.org/١٠,١٠١٦/j.eswa.٢٠١٨,٠٣,٠٢٥

[١٣] Oreski ,S., oreski ,G.,(٢٠١٤). Genetic algorithm-based heuristic for feature selection in credit risk assessment, Expert systems with applications Journal, Vol.٤١,pp.٢٠٥٢-٢٠٦٤.http://dx.doi.org/١٠,١٠١٦/j.eswa.٢٠١٣,٠٩,٠٠٤

[١٤] Tai ,LeQuy. Huyen ,GiangThi Thu. "Deep Learning Techniques for Credit Scoring," Journal of Economics, Business and Management vol. ٧, no. ٣, pp. ٩٣-٩٦, ٢٠١٩.Doi:١٠,١٨١٧٨/joebm.٢٠١٩,٧,٣,٥٨٨

[١٥] Meng, CZ., Liu ,BS., Zhiu ,L. (٢٠١٩). The Practice Study of Consumer Credit Risk Based on Random Forest, Intelligent Systems Research, vol. ١٦٨, pp. ١–٦.

[١٦] Zhao , W., Hou ,J., Ran ,Q. (٢٠٢٢). Analysis of Corporate Credit Risk Based on Random Forest and TOPSIS Models, Financial Engineering and Risk Management, vol. ٥, no. ٤, pp. ٣٠–٣٧.DOI: ١٠,٢٣٩٧٧/ferm.٢٠٢٢,٠٥٠٤٠٥

[١٧] Wang , K., Li ,M., Cheng ,J., Xhou ,X., Li ,G. (٢٠٢٢). Research On personal credit risk evaluation based on XGBoost, Procedia Computer Science, vol. ١٩٩, pp. ١١٢٨–١١٣٥.

[١٨] Li ,Y., Lin ,X., Wang ,X., Shen , F., Gong ,Z.(٢٠١٧) Credit risk assessment Algoithm using deep neural networks with clustering and merging, Vol. ١٣, pp.١٧٣-١٨٤. Doi : ١٠,١١٠٩/CIS.٢٠١٧,٠٠٠٤٥

[١٩] Zhu ,B., Ynag ,W., Wang ,H., Yuan ,Y.(٢٠١٨). A hybrid deep learning model for consumer credit scoring, International conference on artificial intelligence and big data, pp.٢٠٥-٢١٢. Doi : ١٠,١١٠٩/ICAIBD.٢٠١٨,٨٣٩٦١٩٥

[٢٠] Addo ,P.M., Guegan ,D., Hassani ,B.(٢٠١٨). Credit risk analysis using machine and deep learning models, MDPI Journal, Vol.٦, pp.١-١٦.

[٢١] Tyagi , S. (٢٠٢٢). Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions, American International journal of business management, vol. ٥, no. ١, pp. ٥–١٩.

[٢٢] Bastos, J. A. (٢٠٢٢). Predicting Credit Scores with Boosted Decision Trees. Forecasting, ٤(٤), ٩٢٥-٩٣٥. https://doi.org/١٠,٣٣٩٠/forecast٤٠٤٠٠٥٠

[۲۳] Mestiri ,Sami. Hiboun ,Sidi Messaoud. (۲۰۲٤). Credit scoring using machine learning and deep Learning-Based models, Data Science in Finance And Economics,Vol.٤, no.۲.pp.۲۳٦-۲٤۸.DOI:۱۰,۳۹۳٤/DSFE.۲۰۲٤۰۰۹

[۲٤] Jadwal , PK., Agarwal ,S. (۲۰۲۰). Financial credit risk evaluation model using machine learning based approach, Inder Science, vol. ۱٦, no.٦, pp. ۱–۱۲.

[۲٥] Safari, Saeed., Ebrahimi Shaghaghi, Mrziyeh., Sheykh, Mohammad Javad.(۲۰۰۹). Credit risk management of corporate clients in commercial banks with a Data Envelopment Analysis approach (credit rating),Journal of Management Research in Iran,Vol.۱٤,No.۱٤,pp. ۱۳۷-۱٦٤, In Persian.

[۲٦] Safayee Ghadikolayi, Abdolhamid., Ghasemnia Arabi, Narjes.(۲۰۱٦). A new approach to the application of multi-criteria decision-making models in classifying bank credit customers,Modern research in decision making,Vol.۱,No.۳,pp. ٤٤-٦۳, In Persian.

[۲۷] Horri, Mohammad Sadegh., Mahdavi, Kaveh.(۲۰۱٥). Designing a model to predict the credit rating of bank customers using a multi-criteria metaheuristic and hybrid ant colony fuzzy neural network algorithm,Journal of Management Research in Iran,Vol.۱۹,No.۱,pp. ۹۲-۱۱٦, In Persian.

[۲۸] Sohrabi, Babak., Raeesi Vanani, Iman., Zare Mirak Abad, Faeze.(۲۰۱٦). Designing a recommender system to optimize and manage banking facilities based on facility clustering and classification algorithms,Modern research in decision making,Vol.۱,No.۲,pp. ٦۳-۷٦, In Persian.

[۲۹] Alborzi, Mahmoud., Fallah, Mir Feyz., Armaki, Ali.(۲۰۲۲). Development and Explanation of Bank Customer Credit Rating System Based on Hybrid Meta-Learning Models: Case Study, Mellat Bank,Financial Management Perspectives Journal, Vol.۱۲,No.۳,pp. ٦۹-۹٤, In Persian.

[۳۰] Schmid , L., Roidl,M., Pauly ,M. (۲۰۲٤). Comparing statistical and machine learning methods for time series forecasting in data-driven logistics – A simulation study, International journal of Arxiv, vol. ۲, pp. ۱–۳۹.

[۳۱] Brzozowska , j., Pizon ,J., Baytikenova ,G., Gola ,A., Zakimova ,A., Piotrowska ,K. (۲۰۲۳). DATA ENGINEERING IN CRISP-DM PROCESS PRODUCTION DATA – CASE STUDY, Applied Computer Science, vol. ۱۹, no. ۳, pp. ۸۳–۹٥.doi: ۱۰,۳٥۷۸٤/acs-۲۰۲۳-۲٦

[۳۲] Nazari , A., Mehregan ,M., Tehrani,R.(۲۰۱۹). Credit Scoring of Bank Depositor with Clustering Techniques for Supply Chain Finance, International Journal of Supply Chain Management, vol. ۸, no. ۱, pp. ۳۷٤–۳۸۳.

[۳۳] Lenssen , L., Schubert ,E. (۲۰۲۳). Medoid Silhouette clustering with automatic cluster number selection, International journal of Arxiv, vol. ۱, , pp. ۱–۱۳.DOI:۱۰,۱۰۰۷/۹۷۸-۳-۰۳۱-۱۷۸٤۹-۸_۱٥

[۳٤] Deviatiarova , E., Fadeev ,S., Dukhanov ,A. (۲۰۲۳). Analyzing Proficiency Patterns in Test Results Using Clustering and Augmentation Algorithms, Procedia Computer Science, vol. ۲۲۹, , pp. ۲٥٤–۲٦٤.

[۳٥] Sutramiani , NP., Aurelia ,SH., Fauzi ,M. (۲۰۲٤). The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover, Journal of Information Systems Engineering and Business Intelligence, vol. ۱۰, no. ۱, pp. ۱–۱۲.Doi: http://dx.doi.org/۱۰,۲۰٤۷۳/jisebi.۱۰,۱,۱۳-۲٤